Exploring Robust Methods for Testing Equality of Variances

Olumide Adesina
Redeemer's University

Tolulope Femi Adesina (Nee Oladeji)
Covenant University Ota Ogun State, Nigeria

International Journal of Statistics and Economics;
[Formerly known as the "Bulletin of Statistics & Economics" (ISSN 0973-7022)];
ISSN 0975-556X; Year: 2018, Volume: 19, Issue Number: 3; Int. j. stat. econ.;

Exploring Robust Methods for Testing Equality of Variances

Ismaila Adeleke[1], **Tolulope F. Oladeji**[2] and **Olumide S. Adesina**[3]

[1]Department of Actuarial Science & Insurance
University of Lagos, Akoka, Nigeria
adeleke22000@gmail.com

[2]Department of Banking & Finance
Covenant University, Ota Ogun State, Nigeria
oladeji.tolulope@covenantuniversity.edu.ng

[3]Department of Mathematical Sciences
Olabisi Onabanjo University, Nigeria
olumidestats@gmail.com

ABSTRACT

Assessment and test of equality of volatilities of stocks with similar returns is important for decision makers in their portfolio selection. Volatility estimation is equally important for pricing stocks and derivative securities. In this study, three robust tests for homogeneity of variance are considered in the selection of stocks under the conditions of high kurtosis and large skewness. Investigations involved the effectiveness of the Ballett's test, robust methods of Levene's Test and Fligner-Killeen test in the test of homogeneity of variance of stocks. Simulated and insurance stock data were employed and empirical results findings show that the three test statistics are equally efficient in testing homogeneity of variance even as the datasets deviates from normality.

Keywords: Homogeneity, Levene's Test and Fligner-Killeen test, Ballett's test, Normality, Stocks, Test of Equality of Volatility.

Subject Classification Numbers: C58

1. INTRODUCTION

In many statistical applications a test for the equality of variance is of interest; test for equality of variance can be used alongside other methods to support assumptions made about variance (Nordstokke *et. al*., 2011). Homoscedasticity condition is also important in regression analysis as ignoring heteroscedasticity can lead to inefficient estimation or incorrect inference (Ruppert, Wand and Carroll, 2003). However, the test of equality of variance can be used to answer questions of whether two or more samples have equal variance; and determining if a new practice, treatment, or test reduces the variability of the current process. The assumption of equality of variances is based on the premise that the population variances on the variable being analyzed are equal. Nevertheless, validity of the results is jeopardized if the hypothesis of homogeneity of variances is upheld when indeed variances are unequal (Glass *et al*., 1972). It is important for decision makers to use robust methods to assess and test equality of volatility among stocks that have similar returns characteristics as they will be able to make informed decisions in their portfolio selection. When investors are guided in their decisions, they tend to select stocks that will yield better returns as they have been informed about characteristics of such as a result of the use of robust methods. According to Bali, Cakici & Whitelaw (2011), investors may be willing to pay more for stocks that exhibit extreme positive returns. Investors that are well informed in their portfolio selection are also able to reduce or eliminate non compensated risk (Goetzmann & Kumar, 2008). The information that investors are able to gather from the use of robust methods helps them in the effective diversification of their portfolio, as it helps them to avoid under diversification which could be very costly (Goetzmann & Kumar, 2008).

Procedures for comparing variances can also be used as a preliminary test in analysis of variance, dose-response modeling or discriminant analysis. The F-test for equality of variance is used as a statistic for testing variance just as Bartlett's test of equal covariance matrices is used to decide between fitting linear and quadratic discriminant functions (Boos and Brownie 2004). The two sample T-test is a reliable method in testing the differences between populations mean, just as the classical F-test is in testing the differences in population variance, but are sensitive to the assumptions of normality and have problems of too many Type I errors and reduced power. Studies investigating robust test statistics that are appropriate under conditions of non-normality and variance heterogeneity include; Box (1953), Levene (1960), Brown and Forsythe (1974), Box and Anderson (1955), Conover, Johnson and Johnson (1981), Boos and Brownie (2004). For heavily tailed and highly skewed distributions, the Levene (1960) test always has too many rejections. As an improvement to Levene, Brown and Forsythe (1974) introduced a modification that replaces the central location $\overline{y}_i$ by resistant statistics like the median and the trimmed mean. Methods based on Monte Carlo simulations had also been worked on (Conover, Johnson and Johnson, 1981). This methodology identifies procedures that are robust with respect to test size and power over a wide range of distributions and sample size. Oral (2012) derived the modified maximum likelihood estimators of the mean and the standard deviation, the authors also came up with procedures for hypothesis testing of the population mean under non-normality condition and when a dataset does not necessarily follow lognormal population. Covrig *et. al.,* (2011) pointed out that actuary faces the problem of the need to determine the distribution function of a sum of random variables which are not certainly independent in the non-life insurance business.

This study explores various robust tests of variance in the selection of stocks that are similar in returns under the conditions of high kurtosis and large skewness. Effectiveness of the Ballett's test, robust methods of Levene's Test and Fligner-Killeen test in the test of homogeneity of variance of stocks is investigated. The remainder of this paper is arranged thus: in section 2, material and method is highlighted. In section 3, simulation study was carried out from various distributions and discussed. Section 4 contains the statistical findings, and section 5 contains summary and conclusion of the study.

2. MATERIALS AND METHODS

The power of a statistical test is the probability of correctly rejecting a false null hypothesis, that is the probability that the test will conclude that the phenomenon exists (Cohen, 1988). If the power of an experiment is low, then there is a good chance that the experiment will be inconclusive. Hence, power is the ability of a test to correctly reject the null hypothesis. Robustness of validity is the is the ability of a procedure to control the Type I error rate of a test close to the nominal value and stable over a range of distributions even with some deviations from its assumptions (Guo and Luh, 2000; Sawilowsky and Blair, 1992). The study examines equality and homogeneity of variance of four insurance opening stock prices. In test for homogeneity of stock price variance, null hypothesis of homoscedasticity is stated in the form:

$$H_0 : \sigma_1^2 = \sigma_2^2 = ...\sigma_a^2$$
(2.1)

the alternative hypothesis is that at least two of them are heteroscedastic.

Ballett's Test

The algorithm for performing Bartllet's test follows the following steps as given by (Bartllet 1937)

i. Calculate $U = \frac{1}{C}\left[vln(s_p^2) - \sum_{i=1} av_i ln(s_i^2) \right]$ (2.2)

Where,

$$s_p^2 = \frac{\sum_{i=1}^{a} v_i s_i^2}{v}, v_i = n_i - 1, v = \sum_{i=1}^{a} v_i$$
(2.3)

$$C = 1 + \frac{1}{3(a-1)}\left(\sum_{i=1}^{a} \frac{1}{v_i} - \frac{1}{v} \right)$$
(2.4)

For one way ANOVA $s_p^2 = MSE$ and $v = N - a$

ii. Reject $H_0 : \sigma_1^2 = \sigma_2^2 = ...\sigma_a^2$ if $U > \chi^2(\alpha, a-1)$ (2.5)

The Bartlett's Test is the uniformly most powerful (UMP) test for the homogeneity of variances problem under the assumption that each treatment population is normally distributed (Bartllet 1937). Its weakness is in the normality assumption. As a result, many statisticians do not recommend its use, Vorapongsathorn *et. al.,* (2004). The Levene's Test or the Fligner-Killeen Test is usually recommended because they are not very sensitive to departures from normality.

Levene's Test

The algorithm for this test is Levene (1960):
i. Calculate each $z_{ij} = | y_{ij} - \overline{y}_i |$

ii. Run an ANOVA on the set of z_{ij} values

iii. if p-values $\leq \alpha$, reject H_0 and conclude that variances are no equal

Levene's Test is robust because the true significance level is very close to the nominal significance level for a large variety of distributions (Gastwirth et. al 2009) but is not sensitive to symmetric or heavy-tailed distributions NIST (2006).

The Fligner-Killeen Test

Fligner and Killeen (1976) suggested jointly ranking the absolute values $|X_{ij}|$ and assigning increasing scores such as $a_{N,i} = i, a_{N,i} = i^2$ or $a_{N,i} = \Phi^{-1}((1+i/(N+1))/2)$))/2) based on the ranks. Conover, Johnson, and Johnson (1981) modified the Fligner-Killeen test and suggested ranking $|X_{ij} - \tilde{X}_j|$ instead of $|X_{ij}|$, where $\tilde{X}_j$ is the sample median of the j^{th} population. This version of the Fligner-Killeen test is called the median-centering Fligner-Killeen test (Niu 2004).

The modified Fligner-Killeen test can be calculated by

$$\chi^2 = \sum_{j=1}^{k} n_j (\overline{A}_j - \overline{a})^2 / V^2$$

(2.6)

where $\overline{A}_j$ is the mean score for the j^{th} sample, a is the overall mean score , i.e.

$$\overline{a} = \frac{1}{N}\sum_{i=1}^{N} a_{N,i} \text{ , and } V^2 = \frac{1}{N-1}\sum_{i=1}^{N}(a_{N,i} - \overline{a})^2$$

(2.7)

For large sample sizes, the Fligner-Killeen statistic χ^2 has an asymptotic chi-square distribution with $(k-1)$ degrees of freedom.

The stock prices of each insurance company were ranked and grouped as shown in table 1 below, while in table 5, the three tests above were used to test for equality of their variances.

3. SIMULATION STUDY

In this session, results of simulation studies are presented. Various sizes of 10, 20, 30, 50 and 100 samples were drawn from normal, log-normal, skewed-normal, and skewed-t distributions. The samples are grouped based on descriptive statistics for each distribution; therefore, the groupings did not follow the same pattern. Software package by R Core team (2017) was used to implement the analysis. Criteria used to determine performance of each model was their p-values, whereas Kosapattarapim *et. al.,* (2012) used the AIC criterion to determine the best fitting model for simulated data, after which same criterion was applied to stock exchanges.

Table 1: Distribution of simulated samples from Normal Distribution

Rank	Group	*Frequency*	*frequency*	*frequency*	*frequency*	*frequency*
A	-3.5 to -2.0	1	0	0	2	2
B	-2.0 to -0.5	1	5	1	16	29
C	-0.5 to 1.0	2	13	5	24	50
D	1.0 to 2.5	6	2	17	8	18
E	2.5 to 4.0	0	0	0	0	1
Total		10	20	30	50	100

Table 2: Distribution of simulated samples from Log-Normal Distribution

Rank	Group	*frequency*	*frequency*	*frequency*	*frequency*	*frequency*
A	0.05 to 0.1	1	0	0	1	3
B	0.1 to 0.2	1	1	2	1	4
C	0.2 to 1.0	5	10	9	24	37
D	1.0 to 5.0	3	8	17	22	49
E	5.0 to 10	0	1	2	2	6
F	10 to 15	0	0	0	0	1
Total		10	20	30	50	100

Table 3: Distribution of simulated samples from Skewed-Normal Distribution

Rank	Group	*frequency*	*frequency*	*frequency*	*frequency*	*frequency*
A	-0.1 to 0.01	0	2	2	6	8
B	0.1 to 1.0	7	13	17	28	56
C	0.1 to 1.5	2	5	8	8	16
D	1.5 to 2.0	0	0	1	4	11
E	2.0 to 2.5	0	0	0	3	2
F	2.5 to 3.0	0	0	0	1	1
Total		10	20	30	50	100

Table 4: Distribution of simulated samples from Skewed-t Distribution

Rank	Group	*frequency*	*frequency*	*frequency*	*frequency*	*frequency*
A	-1.0×10^{40} to (-1.0×10^{10})	2	3	3	10	20
B	(-1.0×10^{10}) to 1.0	6	10	13	19	30
C	1.0 to (1.0×10^{10})	2	6	11	17	35
D	(1.0×10^{10}) to (1.0×10^{20})	0	2	2	3	11
E	(1.0×10^{20}) to (1.0×10^{30})	0	0	0	1	3
F	(1.0×10^{30}) to (1.0×10^{40})	0	0	0	0	1
Total		10	20	30	50	100

Table 5: Tests of Equality of variance

Normal Distribution	Test	Test Statistics	*df*	F-value	p-value
	Bartlett's χ^2	38.204	4	-	1.017e-07
	Levene's Test	-	4	2.515	0.07389
	Fligner-Killeen χ^2	10.092	4	-	0.00389
Log-Normal	Bartlett's χ^2	53.575	5	-	2.563e-10
	Levene's Test	-	5	2.8455	0.03723
	Fligner-Killeen χ^2	11.978	5	-	0.0351
Skewed-normal	Bartlett's χ^2	47.993	6	-	1.186e-08
	Levene's Test	-	6	2.27569	0.06487
	Fligner-Killeen χ^2	10.886	6	-	0.09196
Skew-t	Bartlett's χ^2	31.389	5	-	7.85e-06
	Levene's Test	-	5	1.7976	0.1515
	Fligner-Killeen χ^2	8.751	5	-	0.1194

4. DATA AND STATISTICAL FINDINGS

The Data

Opening stock price per share in Naira of four Insurance companies was collected from Nigeria Stock Exchange (NSE) for the period February 3, 2014 to June 9, 2015. They are AIICO, Continental, Cornerstone and Prestige Insurance. These companies were selected because of considerable large volume of trade on the stocks and regularity of them being traded on the floor of NSE. The data sets are five working day data of three hundred and fifty two (352) trading days.

Statistical Results

Descriptive statistics, normality tests (Doornik-Hansen test, Shapiro-Wilk W, Lilliefors test and Jarque-Bera test) shows that the data sets does not follow a normal distribution except for Continental Insurance. The Q-Q plots equally show that Continental Insurance follows approximately normal distribution. Table 6 presents a range of statistics for the four stocks, they are: mean variance, skewness and kurtosis. It can be observed that there are similarities across the stocks regarding the mean and the variance of the opening prices. Under the assumption of normality using the standard measures of skewness and kurtosis, all distributions are positively skewed except Continental, indicating that they are non-symmetric. Furthermore, they all exhibit high levels of kurtosis (platykurtic for AIICO, Continental and Prestige, while Leptokurtic for Cornerstone), which indicates that these distributions have other tails than normal distributions. The Doornik-Hansen, Shapiro-Wilk, Lilliefors and

Jarque-Bera tests of normality has also been calculated and presented in Table 7, acceptance or rejection of hypotheses is based on 5% alpha level.

Table 6: Descriptive Statistics of Opening Stock prices of AIICO, Continental, Cornerstone and Prestige Insurance

	Minimum	Maximum	Mean	STDV	Variance	Skewness	Kurtosis
AIICO	.65	1.19	.8424	.08658	.007	1.414	1.859
Continental	.80	1.45	1.1046	.10452	.011	-.030	.246
Cornerstone	.50	.72	.5159	.03629	.001	2.993	9.941
Prestige	.58	1.55	.9436	.18994	.036	1.050	.899

Table 7: Normality Test of Stock prices of AIICO, Continental, Cornerstone and Prestige Insurance

Company	Doornik-Hansen test	Shapiro-Wilk W	Lilliefors test	Jarque-Bera test
AIICO	219.446 (2.22768e-048)	0.847037 (4.88227e-018)	0.212808 (0)	164.589 (1.81943e-036)
Continental	1.40463 (0.495436)	0.99132 (0.0367044)	0.0616428 (0)	0.802654 (0.669431)
Cornerstone	1231.47 (3.88399e-268)	0.510772 (4.84302e-030)	0.390455 (0)	0.390455 (0)
Prestige	106.193 (8.71891e-024)	0.919953 (9.27458e-013)	0.144635 (0)	75.1659 (4.7636e-017)

H_0: Dataset is normally distribution

In table 7, test statistics and (p-value) are presented in each two-line entry

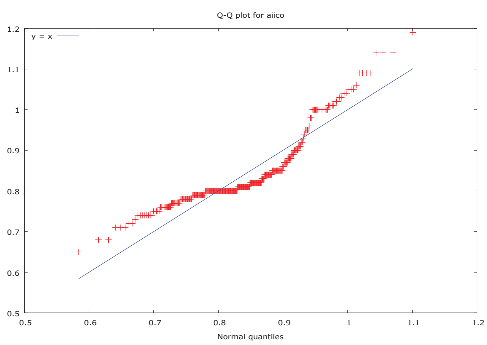

Fig 1: Normal Q-Q Plots for AIICO Insurance

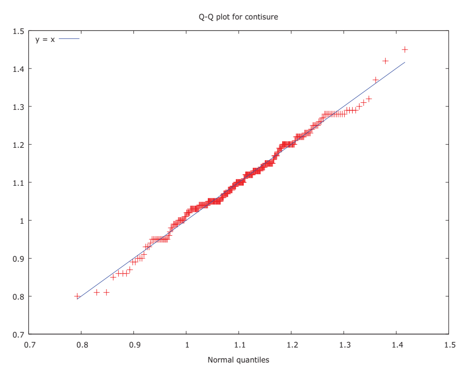

Fig 2: Normal Q-Q Plots for Continental Insurance

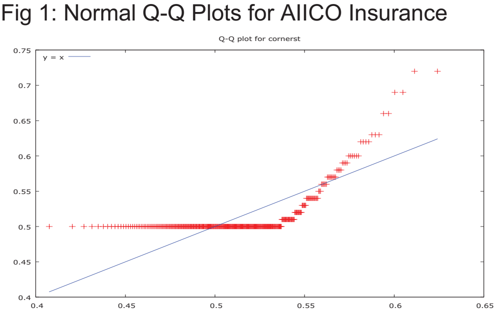

Fig 3: Normal Q-Q Plots for Cornrestone Insurance

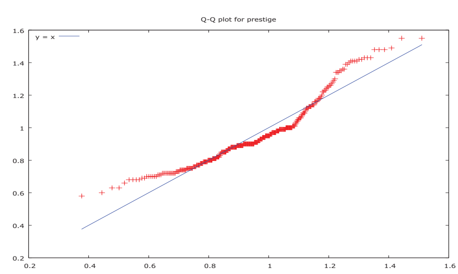

Fig 3: Normal Q-Q Plots for Prestige Insurance

Table 8: Ranking of Stock Prices in naira of AIICO, Continental, Cornerstone and Prestige Insurance

RANK (Naira)	AIICO	Continental	Cornerstone	Prestige
A (0.5-0.6)	0	0	339	2
B (0.6-0.7)	3	0	17	16
C (0.7-0.8)	152	0	2	75
D (0.8-0.9)	209	13	0	105
E (0.9-1.0)	44	43	0	113
F (1.0-1.1)	38	139	0	36
G (1.1-1.2)	4	131	0	24
H (1.2-1.3)	0	75	0	14
I (1.3-1.4)	0	0	0	11
J (1.4-1.5)	0	0	0	14
K (1.6-1.7)	0	0	0	2
Total count	450	401	358	412

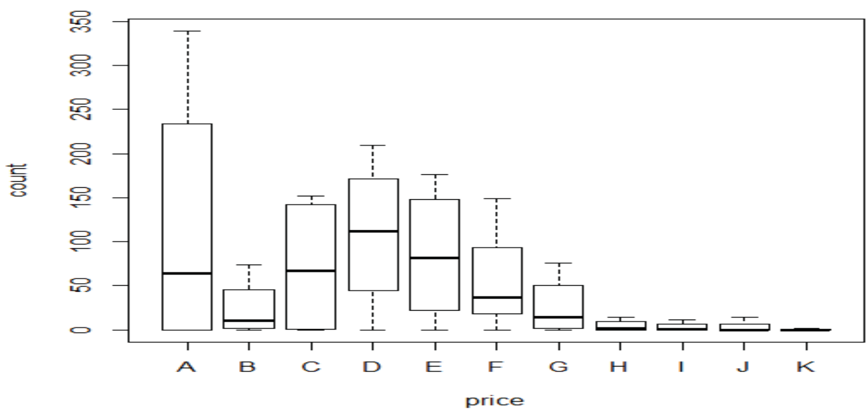

Fig 5: Box-plots of each categories of stock price groupings.

Table 9: Test Statistics for Homogeneity of variance

	Test Statistics	df	F-value	p-value
Bartlett's χ^2	61.808	10	-	1.646e-09
Levene's Test	-	10	3.2768	0.004833
Fligner-Killeen χ^2	26.488	10	-	0.003136

5. SUMMARY AND CONCLUSION

Tables 1-4 displayed the distributions of simulated data from normal, log-normal, skewed-normal, and skewed-t distributions respectively for sample sizes $n = 10$, 20, 30, 50 and 100. Parameters (0, 1) is selected for normal and log-normal, while (0, 1, 5) was used for the Skewed normal and for skewed-t distribution (0.05, 1). The results of tests of normality are presented in Table 5 followed by the Levene and Flinger-Killeen tests of homogeneity. Descriptive statistics of the selected insurance stock prices are displayed in Table 6, followed by tests of normality and tests of homogeneity of variance respectively in Tables 7 and 9. Going by the normality tests and the Q-Q plots, continental insurance stock prices exhibited normality traits contrary to the other insurance stocks that are non-normal in distribution.

In this study, various robust tests of equality of variance are considered in the selection of stocks that are similar in returns under the conditions of high kurtosis and large skewness. This is particularly important for decision makers in the optimum selection of securities in a portfolio. More importantly, every financial analyst use robust methods to access and test equality of volatility among stocks with similar returns expectations in their portfolio selection. The Simulation study shows p-values less than 5 percent alpha level, except for Levene test which is found to be 0.074. For a real-life dataset, the effectiveness of the Ballett's test, robust methods of Levene's Test and Fligner-Killeen test in the test of homogeneity of variance of stocks have been investigated. Findings reveal that the three test statistics are equally efficient in testing homogeneity of variance even when the data set deviates from normality. The results from this study would help investors to identify a statistical test to identify variances, which guides in their decisions on selecting stocks that will yield better returns having in mind the level of risks and volatility inherent in them.

REFERENCES

Bali, T. G., Cakici, N., Whitelaw, R. F. 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. Journal of Financial Economics. 99(2), 427-446.

Bartllet M.S. 1937. Properties of Sufficiency and Statistical Tests, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* Vol. 160, No. 901 268-282

Boos, D. D. and Brownie, C. 2004. Comparing variance and other measures of dispersion.Box. G. E. P. 1953. Non-normality and test on variances. *Biometrika* 40 318-335.

Box.G. E. P. and Anderson, S. L. 1955. Permutation theory in the deviation of robust criteria and the study of departures from assumption (with discussion).*J. Roy. Statist. Soc. Ser. B*17 1-34.

Brown, M. B. and Forsythe, A. B. 1974.Robust test for the equality of variances.*J. Amer*.statist. Assoc. 69 364-367.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. New Jersey: Lawrence Erlbaum Associates.

Conover, W. J., Johnson, M. E. and Johnson, M. M. 1981. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data.*Technometrics* (23) 351–361.

Covrig. M, Mircea. I, Veghes. O and Serban. R . 2011. Some applications of sums of randomvariables in non-life insurance , (8) S11 http://ceser.in/ceserp/index.php/ijms/article/view/2751

Fligner, M.A. & Killen, T.J. 1976. Distribution-free two-sample tests for scale. J. Amer. Statistical Assoc. 71: 210-213.

Gastwirth J. L, Gel Y.R and Weiwen M . 2009. The Impact of Levene's Test of Equality ofVariances on Statistical Theory and Practice. Institute of Mathematical Statistics, Statistical Sciences, Vol. 24, No. 3, 343–360 DOI: 10.1214/09-STS301

Glass, G.V., Peckham, P.D. & Sanders, J.R. 1972. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Review of Educational Research (42) 237-288.

Goetzmann W and Kumar A. 2008. Equity Portfolio Diversification. Review of Finance 12, 433–463.

Guo, J. H., & Luh, W. M. 2000. An invertible transformation two-sample trimmed *t*-statistic under heterogeneity and nonnormality. *Statistic & Probability letters*, *49*, 1-7.

Kosapattarapim. C, Lin. Y and McCrae. M. 2012. Evaluating the Volatility Forecasting Performance of Best Fitting GARCH Models in Emerging Asian Stock Markets. International Journal of

Mathematics & Statistics, 12 (2), 1-15, http://www.ceser.in/ceserp/index.php/ijms/article/view/3198

Layard, M. W. J. 1973. Robust large-sample tests for homogeneity of variances. *J. Amer. Statist. Assoc*. (68) 195–198.

Levene, H. 1960. Robust tests for equality of variances. In *Contributions to Probability and Statistics* (I. Olkin, ed.) 278–292.Stanford Univ. Press, Stanford, CA.

Lix, L. M. &Keselman, H. J. 1998. To trim or not to trim: Tests of mean equality under heteroscedasticity and nonnormality. Educational and Psychological Measurement, 58, 409-429

NIST/SEMATECH. 2006. e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/.

Nordstokke, D. W., Zumbo, B. D., Caims, S. I. and Saklofsk, D. H. 2011. The operating characteristics of the nonparametricsLevene test for equal variances with assessment and evaluation data. *Practical assessment, research and evaluation*, 16(5).

Oral E. 2012. Robust Estimation in Non-normal Samples with Known Detection Limits , Vol 12 Issue 2. http://www.ceser.in/ceserp/index.php/ijms/article/view/3199

R Core Team 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ruppert, D., Wand, M. P. and Carroll, R. J. 2003. Semi-parametric Regression.Electron. J. Statist. Vol 3.1193-1256.

Sawilowsky, S. S., and Blair, R. C. 1992. A more realistic look at the robustness and type II error properties of the t test to departures from population normality.*Psychological Bulletin, 111*, 353-360.

Vorapongsathorn. T , Taejaroenkul. S and Viwatwongkasem. C. 2004. A comparison of type I error and power of Bartlett's test, Levene's test and Cochran's test under violation of assumptions, Songklanakarin J. Sci. Technol. Vol. 26 No. 4

Niu. X. 2004. Statistical Procedures for Testing Homogeneity of Water Quality Parameters. *Department of Statistics Florida State University Tallahassee, FL 32306*